

# Aminoacyl-transferases and the N-end rule pathway of prokaryotic/eukaryotic specificity in a human pathogen

Emmanuelle Graciet<sup>†</sup>, Rong-Gui Hu<sup>†</sup>, Konstantin Piatkov<sup>†</sup>, Joon Haeng Rhee<sup>‡</sup>, Erich M. Schwarz<sup>†</sup>, and Alexander Varshavsky<sup>†§</sup>

<sup>†</sup>Division of Biology, California Institute of Technology, Pasadena, CA 91125; and <sup>‡</sup>National Research Laboratory of Molecular Microbial Pathogenesis and Genome Research Center for Enteropathogenic Bacteria, Chonnam National University Medical School, Gwangju 501-746, Korea

Contributed by Alexander Varshavsky, December 28, 2005

The N-end rule relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. Primary destabilizing N-terminal residues (Nd<sup>P</sup>) are recognized directly by the targeting machinery. The recognition of secondary destabilizing N-terminal residues (Nd<sup>S</sup>) is preceded by conjugation of an Nd<sup>P</sup> residue to Nd<sup>S</sup> of a polypeptide substrate. In eukaryotes, ATE1-encoded arginyl-transferases (R<sup>D,E,C\*</sup>-transferases) conjugate Arg (R), an Nd<sup>P</sup> residue, to Nd<sup>S</sup> residues Asp (D), Glu (E), or oxidized Cys residue (C\*). Ubiquitin ligases recognize the N-terminal Arg of a substrate and target the (ubiquitylated) substrate to the proteasome. In prokaryotes such as *Escherichia coli*, Nd<sup>P</sup> residues Leu (L) or Phe (F) are conjugated, by the *aat*-encoded Leu/Phe-transferase (L/F<sup>K,R</sup>-transferase), to N-terminal Arg or Lys, which are Nd<sup>S</sup> in prokaryotes but Nd<sup>P</sup> in eukaryotes. In prokaryotes, substrates bearing the Nd<sup>P</sup> residues Leu, Phe, Trp, or Tyr are degraded by the proteasome-like ClpAP protease. Despite enzymological similarities between eukaryotic R<sup>D,E,C\*</sup>-transferases and prokaryotic L/F<sup>K,R</sup>-transferases, there is no significant sequelogy (sequence similarity) between them. We identified an aminoacyl-transferase, termed Bpt, in the human pathogen *Vibrio vulnificus*. Although it is a sequelog of eukaryotic R<sup>D,E,C\*</sup>-transferases, this prokaryotic transferase exhibits a "hybrid" specificity, conjugating Nd<sup>P</sup> Leu to Nd<sup>S</sup> Asp or Glu. Another aminoacyl-transferase, termed ATEL1, of the eukaryotic pathogen *Plasmodium falciparum*, is a sequelog of prokaryotic L/F<sup>K,R</sup>-transferases (Aat), but has the specificity of eukaryotic R<sup>D,E,C\*</sup>-transferases (ATE1). Phylogenetic analysis suggests that the substrate specificity of R-transferases arose by two distinct routes during the evolution of eukaryotes.

proteolysis | ClpAP | ClpS | Aat | bacterial protein transferase

A protein substrate of the ubiquitin (Ub) system, which controls the levels of many intracellular proteins, is conjugated to Ub through the action of three enzymes, E1, E2, and E3. The substrate's degradation signal (degron) is recognized by E3 (1–4). A ubiquitylated protein bears a covalently linked poly-Ub chain and is degraded by the 26S proteasome (5, 6). An essential determinant of one class of degrons, called N-degrons, is a substrate's destabilizing N-terminal residue. The set of destabilizing residues in a given cell type yields a rule, called the N-end rule, which relates the *in vivo* half-life of a protein to the identity of its N-terminal residue (7–12). In eukaryotes, the N-degron consists of three main determinants, a destabilizing N-terminal residue of a substrate, its internal Lys (K) residue(s) (the site of formation of a poly-Ub chain), and a conformationally flexible region(s) in the vicinity of these determinants (7, 13, 14).

The N-end rule has a hierarchic structure (Fig. 1). In eukaryotes, N-terminal Asn (N) and Gln (Q) are tertiary destabilizing residues in that they function, through enzymatic deamidation (15, 16), to yield the secondary destabilizing N-terminal residues (Nd<sup>S</sup>) Asp (D) and Glu (E) (Fig. 1A). The activity of Asp and Glu requires their conjugation, by ATE1-encoded isoforms of Arg-tRNA-protein transferase (R<sup>D,E,C\*</sup>-transferase, where C\* is oxidized cysteine

residue), to Arg (R), one of the primary destabilizing N-terminal residues (Nd<sup>P</sup>) (11, 17, 18). The latter are recognized by the E3 Ub ligases of the N-end rule pathway, called N-recognins (Fig. 1A) (7, 9, 10, 12). In mammals, the set of destabilizing residues that function through their arginylation contains not only N-terminal Asp and Glu, but also N-terminal Cys (C), which is arginylated after its nitric oxide (NO)-dependent oxidation (Fig. 1A) (11, 18).

The functions of the eukaryotic N-end rule pathway include the control of peptide import (through the conditional degradation of import's repressor) (8, 9), the fidelity of chromosome segregation (through the degradation of a conditionally produced cohesin's fragment) (19), the regulation of apoptosis (through the degradation of a caspase-processed inhibitor of apoptosis) (20, 21), the regulation of meiosis (10), leaf senescence in plants (22), and cardiovascular development in mammals (11, 18, 23). The latter function is likely to be mediated in part by the arginylation-dependent degradation of RGS4, RGS5, and RGS16, a set of GTPase-activating proteins that down-regulate the signaling by specific G proteins and are themselves down-regulated by the N-end rule pathway, at rates controlled by NO and oxygen (11, 18, 23). The Johanson-Blizzard syndrome, a human genetic disease, is caused by the absence of UBR1, one of the E3 N-recognins of the N-end rule pathway (Fig. 1A) (24).

Although prokaryotes lack Ub conjugation and Ub itself, they contain the (Ub-independent) N-end rule pathway, which has been so far characterized only in *Escherichia coli* (Fig. 1B) (25, 26). The pathway's physiological functions in prokaryotes are unknown. In *E. coli*, the Nd<sup>P</sup> residues Leu (L) or Phe (F) are conjugated, by the *aat*-encoded Leu/Phe-tRNA-protein transferase (L/F<sup>K,R</sup>-transferase), to N-terminal Arg or Lys, which are the Nd<sup>S</sup> residues in prokaryotes but Nd<sup>P</sup> residues in eukaryotes (Fig. 1A and B) (7, 25, 26). Reporter substrates bearing the Nd<sup>P</sup> residues Leu, Phe, Trp (W), or Tyr (Y) are targeted for degradation by ClpAP (27–30), one of several proteasome-like proteases in *E. coli*. ClpS, a 12-kDa adaptor protein specific for ClpAP, is an essential component of the *E. coli* N-end rule pathway, where it functions as the N-recognin (Fig. 1B) (31).

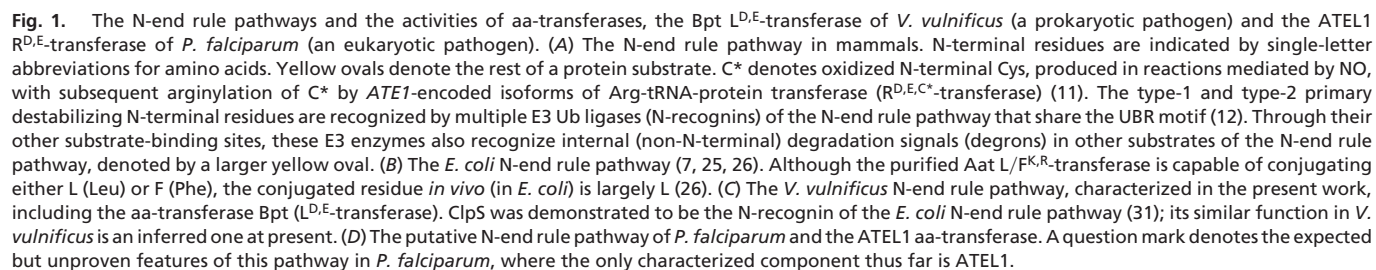
Although eukaryotic R<sup>D,E,C\*</sup>-transferases (ATE1) and prokaryotic L/F<sup>K,R</sup>-transferases (Aat) mediate analogous enzymatic reactions (Fig. 1), there is no significant sequelogy (sequence similarity) between these sets of enzymes. [In this terminology (32), sequelog and spallog denote, respectively, a sequence that is similar, to a specified extent, to another sequence, and a 3D structure that is similar, to a specified extent, to another 3D structure. These terms and their derivatives allow separate, single-word notations for

Conflict of interest statement: No conflicts declared.

Abbreviations: aa-transferase, aminoacyl-transferase; Bpt, bacterial protein transferase; Ub, ubiquitin; Nd<sup>P</sup>, primary destabilizing N-terminal residue; Nd<sup>S</sup>, secondary destabilizing N-terminal residue; C\*, oxidized cysteine residue; βgal, β-galactosidase.

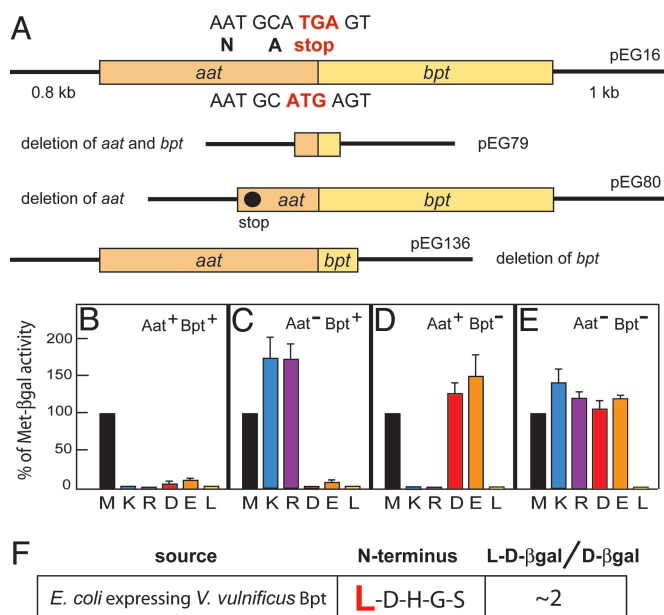
<sup>§</sup>To whom correspondence should be addressed. E-mail: [avarsh@caltech.edu](mailto:avarsh@caltech.edu).

© 2006 by The National Academy of Sciences of the USA



We describe here a prokaryotic aminoacyl-transferase (aa-transferase) termed Bpt (bacterial protein transferase) and characterized in the human pathogen *Vibrio vulnificus*. This Gram-negative  $\gamma$ -proteobacterium is naturally found in marine environments, including shellfish, and can cause gastrointestinal and bloodstream infections in humans. These infections can be severe, and occasionally fatal, in patients with compromised immune system, iron overload, or hepatic disease (ref. 33 and refer-

**An aa-Transferase in *V. vulnificus*.** Through searches in databases we found that some prokaryotes, other than *E. coli*, contain proteins that are sequelogs of eukaryotic ATE1 (see Introduction for the



**Fig. 2.** The N-end rule pathway in *V. vulnificus* and *E. coli*. (A) The *V. vulnificus* *aat-bpt* operon and its derivatives. The ORF encoding the Bpt aa-transferase begins within the stop codon of the preceding *aat* ORF. The nucleotide sequence of the *bpt* ORF (including its start codon, in red) is shown below the diagram. Shown above is the 3' end of *aat* ORF, including its TGA stop codon, in red. Also indicated are the lengths of flanking *V. vulnificus* DNA fragments in the plasmid insert. (B) Relative enzymatic activities of βgal in Aat-lacking *E. coli* (TS351) expressing *S. cerevisiae* UBP1, both aa-transferases of *V. vulnificus* (WT), and either Met-βgal (M), Lys-βgal (K), Arg-βgal (R), Asp-βgal (D), Glu-βgal (E), or Leu-βgal (L), produced from the corresponding Ub fusions. (C) The same as B but in the absence of *V. vulnificus* Aat. (D) The same as B but in the absence of *V. vulnificus* Bpt. (E) The same as B but in the absence of *V. vulnificus* Aat and Bpt. One hundred percent is the (averaged) activity of Met-βgal (M). (F) Determination, through Edman degradation, of the N-terminal sequence of Asp-βgal (derived from Ub-Asp-βgal) from *E. coli* KPS18 that lacked both Aat and ClpA and expressed *V. vulnificus* Bpt. The indicated ratio of Leu-Asp-βgal/Asp-βgal refers to the incomplete *in vivo* leucylation of (overexpressed) Asp-βgal by *V. vulnificus* Bpt in *E. coli*, as detected by Edman sequencing.

terms and notations used). We denoted the corresponding genes as *bpt* to distinguish them from eukaryotic transferase genes (*ATE1*) (11, 18) and prokaryotic *aat* genes (25, 26). Remarkably, the sequenced genome of the human prokaryotic pathogen *V. vulnificus* (NC.004459 and NC.004460) was found to encode Aat (L/F<sup>K,R</sup>-transferase) and Bpt in a two-transferase operon (NC.004459; Vv12124), in which the ORF of *bpt* begins within the stop codon of the preceding *aat* ORF (Fig. 2A). Partial overlaps of adjacent ORFs in an operon were observed with other prokaryotic genes as well (34).

To examine the recognition/conjugation specificity of *V. vulnificus* Bpt, an Aat-lacking strain of *E. coli* (*aat::minitet*, TS351) (26) was transformed with plasmids expressing the yeast deubiquitylating enzyme UBP1 and either the intact *V. vulnificus* *aat-bpt* operon or its derivatives with deletions (Fig. 2A) (see *Materials and Methods* and *Supporting Text*, which is published as supporting information on the PNAS web site). By cleaving, nearly cotranslationally, after the last residue of Ub in a Ub fusion, the UBP1 enzyme enabled the Ub fusion technique in *E. coli* (25, 35), thereby making possible the *in vivo* production of X-β-galactosidase (βgal) reporter proteins in which N-terminal X was a desired residue.

**Bpt Transferase Has a Hybrid Recognition/Conjugation Specificity.** Asp-βgal and Glu-βgal are normally long-lived in *E. coli*, because N-terminal Asp and Glu are stabilizing residues in this prokaryote,

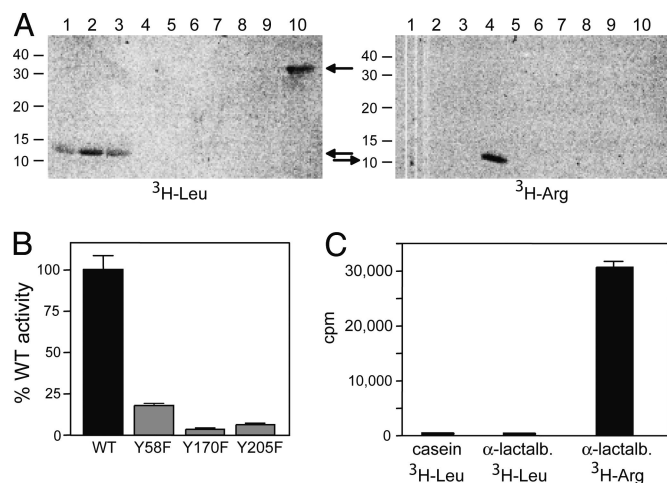
in that they are not recognized by either the Aat L/F<sup>K,R</sup>-transferase or the rest of the *E. coli* N-end rule pathway (25, 26, 31). Previous work (13, 17, 36) has shown that the steady-state level of an X-βgal reporter is a sensitive measure of its metabolic stability. Remarkably, Asp-βgal and Glu-βgal became short-lived in Aat-lacking *E. coli* that expressed *V. vulnificus* Bpt, irrespective of the presence or absence of *V. vulnificus* Aat (Fig. 2B–E). This result indicated that *V. vulnificus* Bpt, similarly to eukaryotic ATE1 (and in contrast to prokaryotic Aat) recognized N-terminal Asp and Glu. This result also indicated that an amino acid conjugated by *V. vulnificus* Bpt could not be Arg, but was expected, instead, to be either Leu, Phe, Trp, or Tyr. The reason for this conclusion is that any one of the latter four residues, if made N-terminal, would be directly recognized as a Nd<sup>P</sup> residue by the ClpS–ClpAP targeting/proteolytic complex (Fig. 1B) (31). Because the degradation of Asp-βgal and Glu-βgal, the substrates of *V. vulnificus* Bpt, was not affected by the absence of *V. vulnificus* Aat, and because a converse omission of *V. vulnificus* Bpt from *E. coli* expressing *V. vulnificus* Aat did not perturb the degradation of Aat substrates Arg-βgal and Lys-βgal (Fig. 2B–E), a physical interaction between Bpt and Aat, even if it normally occurred, would not be required for the activities of these enzymes. Neither coimmunoprecipitation tests nor a cAMP-based two-hybrid *in vivo* assay in *E. coli* (37) suggested a physical interaction between *V. vulnificus* Bpt and Aat (Fig. 5, which is published as supporting information on the PNAS web site, and *Supporting Text*).

To identify directly the amino acid residue(s) conjugated by Bpt to its substrate, Asp-βgal was isolated from *E. coli* (KPS18) that lacked both Aat and ClpA and expressed *V. vulnificus* Bpt, followed by N-terminal (Edman) sequencing. It revealed the N-terminal sequence of leucylated Asp-βgal, Leu-Asp-βgal (Fig. 2F). Thus, *V. vulnificus* Bpt, although it recognizes N-terminal Asp and Glu (similarly to eukaryotic ATE1), conjugates Leu (Nd<sup>P</sup> in both prokaryotes and eukaryotes) to these residues, in contrast to eukaryotic ATE1, which conjugates Arg (Fig. 1A and C). The recognition/conjugation specificity of *V. vulnificus* Bpt was independently confirmed by using *in vitro* conjugation assays with either <sup>3</sup>H-Leu or <sup>3</sup>H-Arg and either α-casein (bearing N-terminal Lys) or bovine α-lactalbumin (bearing N-terminal Glu) as reporters (Fig. 3A). In addition, the enzymatic specificity of strong sequelogs (32) of *V. vulnificus* Bpt from other prokaryotes, such as the plant pathogen *Agrobacterium tumefaciens* and the plant symbiont *Mesorhizobium loti* was found to be indistinguishable from that of *V. vulnificus* Bpt (Fig. 3A). We conclude that Bpt proteins are a distinct class of prokaryotic aa-transferases, termed L<sup>D,E</sup>-transferases, which exhibit a hybrid (prokaryotic/eukaryotic) specificity. It is likely (but remains to be determined) that a Bpt L<sup>D,E</sup>-transferase can also leucylate an oxidized N-terminal Cys, analogously to the arginylation of this residue by R<sup>D,E,C\*</sup>-transferases (Fig. 1A and C) (11, 18).

**Conserved Tyrosines and Cysteines Essential for Bpt Activity.** Sequence alignments of Bpt proteins revealed strong conservation of specific residues (Fig. 6, which is published as supporting information on the PNAS web site, and *Supporting Text*). We mutated the conserved Tyr-58, Tyr-170, and Tyr-205 in *V. vulnificus* Bpt to Phe residues. The resulting Bpt-Y58F, Bpt-Y170F, and Bpt-Y205F were produced in *E. coli* KPS22 (lacking Aat and LacZ) and purified, followed by tests in *in vitro* conjugation assays with <sup>3</sup>H-Leu and α-lactalbumin. Bpt-Y58F lost 70–80% of wild-type Bpt activity (Fig. 3B). Both Bpt-Y170F and Bpt-Y205F, while retaining detectable activity, lost ≈90% of it (Fig. 3B). CD spectra of the mutant proteins were similar to that of wild-type Bpt (data not shown), indicating the absence of strong folding defects. Thus, at least one of the above Tyr residues may be a part of the Bpt's active site.

Mutants of *Saccharomyces cerevisiae* ATE1 that lack N-terminus-proximal Cys residues are inactive *in vitro* (38). Two of these Cys residues of ATE1 were also conserved among Bpt





**Fig. 3.** Enzymatic specificities of *V. vulnificus* Bpt and *P. falciparum* ATEL1 aa-transferases. (A) (Left)  $^3\text{H}$ -Leu: lane 1, purified *V. vulnificus* Bpt was incubated with  $^3\text{H}$ -Leu, other components of the aa-transferase assay, and the 14-kDa bovine  $\alpha$ -lactalbumin (bearing N-terminal Glu) as a reporter. Lanes 2–5, the same as lane 1 but with *A. tumefaciens* Bpt, *M. luti* Bpt, *S. cerevisiae* ATE1, and *E. coli* Aat, respectively. Lanes 6–10, the same as lanes 1–5 but with the 24-kDa bovine  $\alpha$ -casein (bearing N-terminal Lys) as a reporter. (Right)  $^3\text{H}$ -Arg: the same as Left but with  $^3\text{H}$ -Arg instead of  $^3\text{H}$ -Leu. Arrows indicate the bands of  $^3\text{H}$ -lactalbumin and  $^3\text{H}$ -casein, and numbers to the left indicate molecular masses, in kDa, of protein standards. (B) Relative enzymatic activities of *V. vulnificus* Bpt in which specific Tyr residues were converted to Phe. Purified mutant and WT *V. vulnificus* Bpt (0.2  $\mu\text{g}$  each) were assayed *in vitro* with  $^3\text{H}$ -Leu as described in *Supporting Text*, using  $\alpha$ -lactalbumin as a reporter.  $^3\text{H}$  was measured in triplicate samples, and the data were corrected by subtracting  $^3\text{H}$  incorporation in the control assay lacking Bpt. (C) Aa-transferase assay with purified (0.25  $\mu\text{g}$ ) *P. falciparum* ATEL1 enzyme in the presence of either  $\alpha$ -casein or  $\alpha$ -lactalbumin and either  $^3\text{H}$ -Leu or  $^3\text{H}$ -Arg as indicated.

aa-transferases (Fig. 6A). In agreement with *S. cerevisiae* results (38) and the above sequelogy-based deductions, the Cys-18  $\rightarrow$  Ser mutant of *V. vulnificus* Bpt was found to be inactive as well (Fig. 6B). The similarity between prokaryotic Bpt proteins and eukaryotic ATE1 proteins involves a number of other conserved residues as well (Fig. 6A and D).

**ATEL1 of Malaria Parasite Is a Sequel of Prokaryotic Aat Transferases, but Its Specificity Is That of Eukaryotic ATE1 Transferases.** Another kind of aa-transferase identified in the present work is encoded by the genome of *P. falciparum*. This eukaryote is an obligatory intracellular parasite and the cause of malaria in humans (39). BLASTP with *E. coli* Aat as a query predicted, with a moderate but statistically significant *E*-value of  $7\text{e-}4$ , that *P. falciparum* may encode an aa-transferase (NP\_473045.1) that is sequelogous to prokaryotic Aat transferases (L/F<sup>K,R</sup>-transferases). In contrast, the sequenced genomes of mammalian and arthropod species (including the hosts of *P. falciparum*) lack significant sequelogs of Aat (data not shown). Sequelogs of prokaryotic Aat are also present in other *Plasmodium* species (*P. yoelii yoelii*, *P. chabaudi*, and *P. berghei*), with comparable but stronger *E*-values ( $4\text{e-}7$ ,  $1\text{e-}6$  and  $2\text{e-}6$ , respectively, using *E. coli* Aat as a query).

To determine the conjugation/recognition specificity of the putative aa-transferase of *P. falciparum*, it was produced in *E. coli* (lacking aat) and purified, followed by *in vitro* conjugation assays with either  $^3\text{H}$ -Leu or  $^3\text{H}$ -Arg and appropriate reporters. The specificity of the *P. falciparum* aa-transferase was found to be indistinguishable from that of “canonical” eukaryotic ATE1s (Figs. 1 and 3C), despite the absence of significant sequelogy to ATE1s. The *P. falciparum* aa-transferase is thus a R<sup>E</sup>-transferase. We denoted it, and its strong sequelogs in other species, as ATEL1

(“ATE-like”), to distinguish this group of R-transferases from ATE1s and their strong sequelogs Bpt aa-transferases. *In vivo* degradation assays with X- $\beta$ gal reporters in *E. coli* expressing *P. falciparum* ATEL1 confirmed its specificity for N-terminal Glu, and in addition indicated that it arginylates N-terminal Asp (data not shown). Thus, *P. falciparum* ATEL1 is a R<sup>D,E</sup>-transferase (Fig. 1D).

**Identical Specificities of ATE1 and ATEL1 May Be the Result of Convergent Evolution.** PSI-BLAST with *S. cerevisiae* ATE1 as a query did not retrieve sequelogs of ATE1 in *Plasmodium* and related genomes. A converse PSI-BLAST, with *P. falciparum* ATEL1 as a query, did not retrieve sequelogs in eukaryotic genomes known to encode ATE1 enzymes. Strong sequelogs of Bpt L<sup>D,E</sup>-transferases were found only in prokaryotes, where Bpt proteins are frequent in proteobacteria, except for enterobacteria and other prokaryotes that have undergone genome reduction as a result of their obligatory parasitism or symbiosis. Eukaryotes are the only organisms that contain strong sequelogs of mouse/yeast ATE1s, in that prokaryotic Bpt aa-transferases (L<sup>D,E</sup>-transferases) were much less sequelogous (32) to ATE1s of eukaryotes than the latter were among themselves, in addition to Bpt proteins being  $\approx 2$ -fold smaller than eukaryotic ATE1s (for details, see the legend to Fig. 7, which is published as supporting information on the PNAS web site).

To clarify evolutionary links among aa-transferases, we carried out phylogenetic analyses of ATE1, ATEL1, Aat, and Bpt (Figs. 4 and 7 and *Supporting Text*), using both parsimony and distance analyses. The resulting phylogenetic trees were consistent with each other (Figs. 4 and 7). We also tested the Aat/ATEL1 family and the Bpt/ATE1 family for sequelogy (32) to each other (Fig. 6), because even marginally significant sequelogy between them would be consistent with their (remote) homology. Prokaryotic Bpt L<sup>D,E</sup>-transferases and eukaryotic ATE1 R<sup>D,E,C\*</sup>-transferases (Fig. 1) were clearly sequelogous (Fig. 6A and D), strongly suggesting their homology. However, we could not detect sequelogy between prokaryotic Aat L/F<sup>K,R</sup>-transferases and Bpt L<sup>D,E</sup>-transferases (data not shown), even though both Aat and Bpt conjugate Leu to their substrates’ N-terminal residues (Fig. 1B and C). In contrast, the prokaryotic/eukaryotic Aat/ATEL1 aa-transferases and, separately, the prokaryotic/eukaryotic Bpt/ATE1 aa-transferases form two strongly coherent phylogenetic groups within each protein family (Figs. 4 and 7). Thus, phylogenetic analyses suggest that ancestor proteins of *P. falciparum* (eukaryotic) ATEL1 were derived from prokaryotic Aat aa-transferases rather than from ancestors of eukaryotic ATE1s, and that the identical substrate specificities of the contemporary ATE1s and ATEL1s (Fig. 1A and D) may be the result of convergent evolution. Examples of such evolution, where nonsequelogenous, nonspalogous (32) enzymes (see Introduction for definitions of terms) have similar enzymatic properties and physiological functions, include arsenate reductases (40) and vitamin B<sub>6</sub>-dependent enzymes (41).

The distribution of Aat/ATEL1 and Bpt/ATE1 proteins in extant organisms is striking in its unevenness. Although many proteobacterial genomes encode both Aat and Bpt aa-transferases, other bacteria have only Aat or only Bpt, and no detectable Aat or Bpt exist in archaea. ATE1 genes are present in a broad set of “better-known” eukaryotes, such as fungi, plants, animals, and protists. However, we found that ATEL1-like proteins, while apparently absent from the above eukaryotes, are also distributed broadly among other eukaryotes, such as apicomplexans, leishmanias, trypanosomes, cnidaria, diatoms, dinoflagellates, and oomycetes (see Fig. 7 for details). These patterns are consistent with the hypothesis that aa-transferases arose twice, independently, in a common ancestor of proteobacteria and eukaryotes, and underwent two kinds of evolutionary changes in the eukaryotic lineage. First, ATE1 and ATEL1 independently evolved to function as R-transferases (rather than L-transferases, their presumed ancestral states). Second, either ATE1 or ATEL1 were lost from various





(a protein containing the Ub fold) emerged (before the appearance of eukaryotes) as an N-terminal cotranslational chaperone of very early (suboptimally folding) proteins, playing a role analogous to the one that modern Ub appears to play as a cotranslational chaperone of two specific ribosomal proteins (48). The chaperone function of primordial Ub might account for its initial spread through positive selection, before the appearance of Ub-specific enzymes. Apart from the known similarities between the eukaryotic 26S proteasome and its prokaryotic counterparts such as ClpAP (5, 27, 28, 30), the functional analogy (as well as sequeology) between the prokaryotic N-recogin ClpS and the N-recognins (Ub ligases) of the eukaryotic N-end rule pathway (31) is unique so far, to the best of our knowledge, in linking the processive proteolysis in prokaryotes and the Ub system in eukaryotes.

## Materials and Methods

**Construction of *E. coli* Mutants.** The strains and plasmids used, as well as their construction, are described in *Supporting Text* and Tables 1 and 2, which are published as supporting information on the PNAS web site.

**Cloning and Alterations of the *V. vulnificus* *aat-bpt* Operon.** The plasmid pEG16 (which contained the *aat-bpt* operon) and its deletion derivatives were constructed as described in *Supporting Text*.

**Plasmids Expressing X- $\beta$ gal Reporters and UBPI.** pUB23-X plasmids encoding Ub-X- $\beta$ gal fusion proteins (35, 49) were converted into plasmids that coexpressed these fusions and the *S. cerevisiae* UBPI deubiquitylating enzyme, as described in *Supporting Text*.

**Assays for  $\beta$ gal Activity and N-Terminal Sequencing of X.** These procedures were carried out as described in *Supporting Text*.

**Expression and Purification of aa-Transferases.** See *Supporting Text* for description of the cloning of the *E. coli* *aat*, *V. vulnificus* *bpt*,

*M. loti* *bpt*, *A. tumefaciens* *bpt*, and *P. falciparum* *ATEL1* genes, their expression in specific *E. coli* strains, and purification of the corresponding enzymes.

**Site-Directed Mutagenesis of *V. vulnificus* Bpt.** Mutated alleles of the *V. vulnificus* *bpt* gene were obtained as described in *Supporting Text*, followed by expression and purification of the mutant Bpt proteins.

**In Vitro aa-Transferase Assays.** The reporter substrates were either  $\alpha$ -casein (with N-terminal Lys, a substrate of *E. coli* Aat) or bovine  $\alpha$ -lactalbumin (with N-terminal Glu, a substrate of *S. cerevisiae* ATE1). Specific aa-transferases were incubated with substrates in the presence of either  $^3\text{H}$ -Leu or  $^3\text{H}$ -Arg and components (including ATP, tRNAs and aminoacyl-tRNA synthetases) that enabled the amino acid-protein conjugation, followed by SDS/PAGE and autoradiography.

**Phylogenetic Analysis.** Searches for sequelogs (32) of specific aa-transferases were carried out with BLASTP, TBLASTN, or PSI-BLAST. Multiple sequence alignments were produced by using PROBCONS. Phylogenetic analyses were carried out with PHYLIP. See *Supporting Text* for references and details.

We thank J. D. Oliver (University of North Carolina, Charlotte) for the *V. vulnificus* C7184 translucent strain, F. Wellmer (California Institute of Technology) for the *A. tumefaciens* strain C58, T. Kaneko (Kazusa DNA Research Institute, Chiba, Japan) for genomic DNA of *M. loti*, L.-I. Hor (National Cheng-Kung University, Taiwan) for the plasmid pJRD215, the Malaria Research and Reference Reagent Resource Center (MR4; Manassas, VA) for *P. falciparum* 3D7 genomic DNA, which was contributed to MR4 by D. J. Carucci (National Institutes of Health, Bethesda), and D. Ladant (Institut Pasteur, Paris) for the *E. coli*-based two-hybrid system. This study was supported by National Institutes of Health Grants DK39520 and GM31530 (to A.V.). E.G. and K.P. were supported, respectively, by the International Human Frontier Science Program and Colvin postdoctoral fellowships.

- Varshavsky, A. (2005) *Trends Biochem. Sci.* **30**, 283–286.
- Hershko, A., Ciechanover, A. & Varshavsky, A. (2000) *Nat. Med.* **10**, 1073–1081.
- Pickart, C. (2004) *Cell* **116**, 181–190.
- Petroski, M. D. & Deshaies, R. J. (2005) *Nat. Rev. Mol. Cell Biol.* **6**, 9–20.
- Baumeister, W., Walz, J., Zühl, F. & Seemüller, E. (1998) *Cell* **92**, 367–380.
- Wolf, D. H. & Hilt, W. (2004) *Biochim. Biophys. Acta* **1695**, 19–31.
- Varshavsky, A. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 12142–12149.
- Turner, G. C., Du, F. & Varshavsky, A. (2000) *Nature* **405**, 579–583.
- Du, F., Navarro-Garcia, F., Xia, Z., Tasaki, T. & Varshavsky, A. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 14110–14115.
- Kwon, Y. T., Xia, Z. X., An, J. Y., Davydov, I. V., Seo, J. W., Xie, Y. & Varshavsky, A. (2003) *Mol. Cell Biol.* **23**, 8255–8271.
- Hu, R. G., Sheng, J., Xin, Q., Xu, Z., Takahashi, T. T. & Varshavsky, A. (2005) *Nature* **437**, 981–986.
- Tasaki, T., Mulder, L. C. F., Iwamatsu, A., Lee, M. J., Davydov, I. V., Varshavsky, A., Muesing, M. & Kwon, Y. T. (2005) *Mol. Cell Biol.* **25**, 7120–7136.
- Suzuki, T. & Varshavsky, A. (1999) *EMBO J.* **18**, 6017–6026.
- Prakash, S., Tian, L., Ratliff, K. S., Lehotzky, R. E. & Matouschek, A. (2004) *Nat. Struct. Mol. Biol.* **11**, 830–837.
- Baker, R. T. & Varshavsky, A. (1995) *J. Biol. Chem.* **270**, 12065–12074.
- Kwon, Y. T., Balogh, S. A., Davydov, I. V., Kashina, A. S., Yoon, J. K., Xie, Y., Gaur, A., Hyde, L., Denenberg, V. H. & Varshavsky, A. (2000) *Mol. Cell Biol.* **20**, 4135–4148.
- Kwon, Y. T., Kashina, A. S. & Varshavsky, A. (1999) *Mol. Cell Biol.* **19**, 182–193.
- Kwon, Y. T., Kashina, A. S., Davydov, I. V., Hu, R.-G., An, J. Y., Seo, J. W., Du, F. & Varshavsky, A. (2002) *Science* **297**, 96–99.
- Rao, H., Uhlmann, F., Nasmyth, K. & Varshavsky, A. (2001) *Nature* **410**, 955–960.
- Ditzel, M., Wilson, R., Tenev, T., Zachariou, A., Paul, A., Deas, E. & Meier, P. (2003) *Nat. Cell Biol.* **5**, 467–473.
- Varshavsky, A. (2003) *Nat. Cell Biol.* **5**, 373–376.
- Yoshida, S., Ito, M., Gallis, J., Nishida, I. & Watanabe, A. (2002) *Plant J.* **32**, 129–137.
- Lee, M. J., Tasaki, T., Moroi, K., An, J. Y., Kimura, S., Davydov, I. V. & Kwon, Y. T. (2005) *Proc. Natl. Acad. Sci. USA* **102**, 15030–15035.
- Zenker, M., Mayerle, J., Lerch, M. M., Tagariello, A., Zerres, K., Durie, P. R., Beier, M., Hülskamp, G., Guzman, C., Rehder, H., et al. (2005) *Nat. Genet.* **37**, 1345–1350.
- Tobias, J. W., Shrader, T. E., Rocap, G. & Varshavsky, A. (1991) *Science* **254**, 1374–1377.
- Shrader, T. E., Tobias, J. W. & Varshavsky, A. (1993) *J. Bacteriol.* **175**, 4364–4374.
- Gottesman, S. (2003) *Annu. Rev. Cell Dev. Biol.* **19**, 565–587.
- Piszczek, G., Rozycki, J., Singh, S. K., Ginsburg, A. & Maurizi, M. R. (2005) *J. Biol. Chem.* **280**, 12221–12230.
- Burton, B. M. & Baker, T. A. (2005) *Protein Sci.* **14**, 1945–1954.
- Martin, A., Baker, T. A. & Sauer, R. T. (2005) *Nature* **437**, 1115–1120.
- Erbse, A., Schmidt, R., Bornemann, T., Schneider-Mergener, J., Mogk, A., Zahn, R., Dougan, D. A. & Bukau, B. (2006) *Nature* **439**, 753–756.
- Varshavsky, A. (2004) *Curr. Biol.* **14**, R181–R183.
- Kim, Y. R., Kim, S. Y., Kim, C. M., Lee, S. E. & Rhee, J. H. (2005) *FEMS Microbiol. Lett.* **243**, 497–503.
- Fukuda, Y., Nakayama, Y. & Tomita, M. (2004) *Gene* **323**, 181–187.
- Varshavsky, A. (2005) *Methods Enzymol.* **399**, 777–799.
- Xie, Y. & Varshavsky, A. (1999) *EMBO J.* **18**, 6832–6844.
- Karimova, G., Pidoux, J., Ullman, A. & Ladant, D. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 5752–5756.
- Li, J. & Pickart, C. M. (1995) *Biochemistry* **34**, 15829–15837.
- Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, E., Bowman, S., et al. (2002) *Nature* **419**, 498–511.
- Mukhopadhyay, R., Rosen, B. P., Phung, L. T. & Silver, S. (2002) *FEMS Microbiol. Rev.* **26**, 311–325.
- Paiardini, A., Contestabile, R., D'Aguanno, S., Pascarella, S. & Bossa, F. (2003) *Biochim. Biophys. Acta* **1647**, 214–219.
- Uhlmann, F., Lottspeich, F. & Nasmyth, K. (1999) *Nature* **400**, 37–42.
- Zaslöff, M. (2002) *Nature* **415**, 389–395.
- Selitrennikoff, C. P. (2001) *Appl. Environ. Microbiol.* **67**, 2883–2894.
- Lupas, A. N. & Koretke, K. K. (2003) *J. Struct. Biol.* **141**, 77–83.
- Xie, Y. & Varshavsky, A. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 2497–2502.
- Rudolph, M. J., Wuebbens, M. M., Rajagopalan, K. V. & Schindelin, H. (2001) *Nat. Struct. Biol.* **8**, 42–46.
- Finley, D., Bartel, B. & Varshavsky, A. (1989) *Nature* **338**, 394–401.
- Bachmair, A., Finley, D. & Varshavsky, A. (1986) *Science* **234**, 179–186.